A Program for Implementing a Predictive Sample Reuse Approach to Model Selection

by

Shen Yen Lin

University of Minnesota*

Technical Report No. 440

July 1984

KEY WORDS: Predictive Sample Reuse (PSR); Quasi-Bayes Criterion; Quasi-Likelihood Criterion; Nested Models; Likelihood Ratio Test (LRT).

*Supported in part by NIH Grant GM25271

1. Introduction

We wish to choose one of the following models $M_1, M_2, \ldots, M_m$ which will yield the "best" predictions for future observations from the process which generated the given set of data. If the process generating the observations is random and the distribution function $F_k(\cdot|M_k)$ of each of the models is completly specified and known, then the likelihoods under alternative models can be ranked and we can choose the most likely model. In many applied problems, however, we have to consider that the assumptions on the models as less than completely specified.

Predictive sample reuse (PSR) techniques (Geisser 1975) can be applied to High Structure (where distributional forms are assumed known) Selection by Geisser & Eddy (1979). In particular it is assumed that sampling distributions with unknown parameters are given and informative subjective prior distributions for the parameters may not be available. Certain improper prior distributions are used to make up for this lack. The criterion is to maximize the product of conditional predictive densities rather than the maximization of the joint predictive density. The above is given in Sec. 2.

Furthermore, without distributional assumptions, Geisser (1975) gave simple data-analytic solutions, termed low Structure Selections, given in Sec. 3. A computer program for simple regression cases is described in Sec. 4.

2. High Structure Selection

2.1 General Framework

Let $Y = (y_1,\ldots,y_n)$ be observations on N independent random variables $Y_1, Y_2,\ldots,Y_N$.

Assume 1. There is a set of covariates X_j associated with each observation y_j.

2. Several possible models $M_1,\ldots,M_m$, which either partially or completely specify the distributions of Y_j, could offer a satisfactory explanation of the data.

Then, the likelihood, for independent Y_j,

$$L_k = f(Y|X,M_k) = \prod_{j=1}^{N} f_j(y_j|X_j,M_k)$$

where $X = (X_1,\ldots,X_N)$, is a criterion to access the comparisons among the models.

The model M_k specifies the distribution of the Y_j by a set of unknown parameters θ_k. The estimate $\hat{L}_k$ of L_k is made by substituting for θ_k its maximum likelihood estimate (MLE) $\hat{\theta}_k = \hat{\theta}(Y)$ under M_k. The quotient of the two maximized likelihood is the basis for likelihood ratio test (LRT).

Let $Y_{(j)} = (y_1,\ldots,y_{j-1}, y_{j+1},\ldots,y_N)$ be the data set Y with the j^{th} observation Y_j omitted. Let $f(y|Y,X,M_k)$ be a predicting density for future observations when M_k is true. One particular method for choosing the predicting density is discussed. Modify the given density so that $f_j(y_j|Y_{(j)},X_j,M_k)$ is a predicting density for y_j, then

$$L_k = \prod_{j=1}^{N} f_j(Y_j | Y_{(j)}, X_j, M_k), \quad k = 1, \ldots, m .$$

Let $L_k^* = \max_{1 \le k \le m} L_k$, then M_k^* is selected as the best among the appropriate models. Geisser (1966) concluded if the Bayesian Predictive density is used, then the joint distribution of Y_j's is predictively dependent - unconditionally. In L_k, however, Y_j's are treated as predictively independent when conditioned on the parameters.

If we substitute $\hat{\theta}_k$ = the MLE of θ_k, for θ_k in the sampling density $f(y|\theta_k, X, M_k)$, then $f(y|\hat{\theta}_k, X, M_k)$ is termed the quasi-likelihood predicting density. Geisser (1971) suggested another approach,

$$f(y|Y, X, M_k) \propto \int f(y|\theta_k, X, M_k) \prod_{j=1}^{N} f(y_j|\theta_k, X_j, M_k)\, p(\theta_k) d\theta_k$$

where $p(\theta_k)$ is chosen to de-emphasize the effect of the prior in relation to the likelihood and to avoid the introduction of hyperparameters.

2.2 The asymptotic properties of PSR criterion.

Stone (1977) showed that the PSR criterion based on the quasi-likelihood predicting density is asymptotically equivalent to the Akaike (1973) information criterion assuming the model is correct. Geisser and Eddy (1979) stated that in particular if the models are nested with $M_k \subseteq M_k'$ and if the quasi-likelihood predicting density under M_t is

$$f_j(y_j | X_j, \hat{\theta}_{t(j)}, M_t)$$

where $\hat{\theta}_{t(j)}$ is the MLE of θ_t with y_j omitted, then if M_k is true, the ratio

$$\frac{\hat{L}_k}{\hat{L}_{k'}} = \prod_{j=1}^{N} f_j(y_j|X_j,\hat{\theta}_{k(j)},M_k) \Big/ \prod_{j=1}^{N} f_j(y_j|X_j,\hat{\theta}_{k'(j)},M_{k'})$$

converges to

$$\exp\ [-p(M_k) + p(M_{k'}) + \log\lambda] \text{ if } N \to \infty,$$

where $P(M_t)$ is the number of unknown parameters in θ_t. Furthermore,

$$\lambda = \prod_{j=1}^{N} f_j(y_j|X_j,\hat{\theta}_k,M_k) \Big/ \prod_{j=1}^{N} f_j(y_j|X_j,\hat{\theta}_{k'},M_{k'})$$

is LRT criterion.

Under some general conditions, $-2\log\lambda$ is asymptotic χ^2 with $p = p(M_{k'}) - p(M_k)$ d.f.. If considering

H_0: M_k is true

vs.

H_A: $M_{k'}$ is true,

the quasi-likelihood criterion has asymptotic significance level of $\alpha = Pr\ (\chi_p^2 > 2p)$. If the density is exponential or normal, then it will preserve the above asymptotic properties. Schwarz (1978) argued that Akaike's is not consistent and proposed a consistent method. From the point of view of prediction this is of no consequence in nested situations as Geisser and Eddy (1979) point out, see also Clayton, Geisser and Jennings (1984) for numerical studies bearing this out.

2.3 Normal Models

Here PSR is applied to normal models. For exponential models, see Geisser and Eddy (1979).

Assume that the data are samples from one of two normal populations with density

$$f(y|\mu_i,\sigma_i^2) = (2\pi\sigma_i^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma_i^2}(y-\mu_i)^2\right), \quad i=1,2 .$$

Given the model M_1: $u_1 = u_2$, $\sigma_1^2 = \sigma_2^2$, using the usual improper prior density $g(u_i,\sigma_i) \propto \sigma_i^{-1}$, the posterior density of u_i and σ_i is

$$g(u_i,\sigma_i|\bar{y}_i,s_i) \propto f(\bar{y}_i,s_i|u_i,\sigma_i)\cdot\frac{1}{\sigma_i}, \quad i=1,2 ;$$

where

$$s_i^2 = (N_i-1)^{-1}\sum_j (y_{ij}-\bar{y}_i)^2$$

Thus, $$f(y|Y,i,M_1) = \int\int f(y|u_i,\sigma_i)\ g(u_i,\sigma_i|\bar{y}_i,s_i)\,du_i\,d\sigma_i$$

$$= \left[\frac{N}{(N^2-1)\pi}\right]^{1/2} \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{(N-1)}{2}\right)s_i} \left(1+\frac{N(\bar{y}_i-y)^2}{(N^2-1)s_i^2}\right)^{-\frac{N_i}{2}}$$

under our model

$$\propto \left(1+\frac{N(y-\bar{y})^2}{(N^2-1)t^2}\right)^{-\frac{N}{2}} ,$$

where $t^2 = (N-1)^{-1}\sum_i\sum_j (y_{ij}-\bar{y})^2$.

Similarly, for M_2: $u_1 \neq u_2$, $\sigma_1^2 = \sigma_2^2$ and

M_3: $u_1 \neq u_2$, $\sigma_1^2 \neq \sigma_2^2$,

We get, respectively,

$$f(y|Y,i,M_2) \propto \left(1+\frac{N_i(y-\bar{y}_i)^2}{(N_i+1)(N-2)s^2}\right)^{-(N-1)/2}$$

and

$$f(y|Y,i,M_3) \propto \left(1 + \frac{N_i(y-\bar{y}_i)^2}{(N_i^2-1)s_i^2}\right)^{-N_i/2}$$

$$, \quad i=1,2;$$

where $\quad s^2 = (N-2)^{-1} \sum_i (N_i - 1)s_i^2 \quad .$

Each time one observation is omitted and the PSR Quasi-Bayes criterion is selected the model having the largest of L_1, L_2, L_3, where

$$L_1 = \prod_{i=1}^{2} \prod_{j=1}^{N_i} \left[\frac{N-1}{\pi(N-2)N}\right]^{1/2} \frac{\Gamma[(N-1)/2]}{\Gamma[(N-2)/2]t_{(ij)}}$$

$$\left[1 + \frac{(N-1)(y_{ij}-\bar{y}_{(ij)})^2}{N(N-2)t^2_{(ij)}}\right]^{-(N-2)/2}$$

where

$$\bar{y}_{(ij)} = (N-1)^{-1} \sum_{k,t}^{(ij)} y_{kt}$$

$$t^2_{(ij)} = (N-2)^{-1} \sum_{k,t}^{(ij)} (y_{kt} - \bar{y}_{(ij)})^2$$

and

$\sum_{k,t}^{(ij)}$ represents the sum over all values of k,t except (ij) ;

$$L_2 = \prod_{i=1}^{2} \prod_{j=1}^{N_i} \left[\frac{N_i - 1}{\pi(N-3)N_i}\right]^{1/2} \frac{\Gamma[(N-2)/2]}{\Gamma[(N-3)/2]s_{(ij)}}$$

$$\left[1 + \frac{(N_i - 1)(y_{ij} - \bar{y}_{i(j)})^2}{N_i(N-3)s_{(ij)}^2}\right]^{-(N-2)/2},$$

where

$$\bar{y}_{i(j)} = (N_i - 1)^{-1} \sum_t^{(j)} y_{ij}$$

$$s^2_{(ij)} = (N-3)^{-1}\left[(N_i - 2)s_{i(j)}^2 + (N_{3-i} - 1)s_{3-i}^2\right]$$

$$s_{i(j)}^2 = (N_i - 2)^{-1} \sum_t^{(j)} (y_{it} - \bar{y}_{i(j)})^2$$

and $\sum_t^{(j)}$ represents the sum over all values of t except j; and

$$L_3 = \prod_{i=1}^{2} \prod_{j=1}^{N_i} \left[\frac{N_i - 1}{\pi(N_i - 2)N_i}\right]^{1/2} \frac{\Gamma[(N_i - 1)/2}{\Gamma[(N_i - 2)/2]s_{i(j)}}$$

$$\left[1 + \frac{(N_i - 1)(y_{ij} - \bar{y}_{i(j)})^2}{N_i(N_i - 2)s_{i(j)}^2}\right]^{-(N_i - 1)/2}.$$

For the multivariate cases, see Geisser (1964).

2.4 Multiple Regression Problems

Consider the linear model, Y_j independent $\sim N(X_j'\beta, \sigma^2)$, $j = 1,\ldots N$, where $X_j' = (x_{ij},\ldots,x_{qj})$ and $\beta' = (\beta_1,\ldots,\beta_q)$. Geisser (1965) assumed an invariant prior density of the form $g(\beta,\sigma) \propto \frac{1}{\sigma}$ for the predictive density of a future observation. The PSR quasi-Bayes criterion is to choose the largest of L_k where for a subset of arbitrary size k (without loss of generality the first k) of the q predicator variables

$$L_k = \prod_{j=1}^{N} \left(\frac{c_j}{\pi a_{(j)}^2} \right)^{1/2} \frac{\Gamma[(N-K)/2]}{\Gamma[(N-K-1)/2]}$$

$$\left[1 + \frac{c_j (y_j - X_j' b_{(j)})^2}{a_{(j)}^2} \right]^{-(N-K)/2}$$

with

$$c_j = 1 - X_j' (XX')^{-1} X_j$$

$$a_{(j)}^2 = (Y_{(j)} - X_{(j)}' b_{(j)})' (Y_{(j)} - X_{(j)}' b_{(j)})$$

$$b_{(j)} = (X_{(j)} X_{(j)}')^{-1} X_{(j)} Y_{(j)}$$

$$X_{kxN} = (X_1, \ldots, X_N) \quad .$$

$$X_{(j)} = (X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_N)$$

$$Y_{(j)}' = (y_1, \ldots, y_{j-1}, y_{j+1}, \ldots, y_N)$$

$$X_j' = (x_{ij}, \ldots, x_{kj}) \quad .$$

If the order of the variables is rearranged we can get any subset of the model and all possible subsets are compared. The program based on the above formula is given in Sec. 4. Geisser (1965) generalized the above procedure to the normal multivariate regression case.

Example: Hald data from Draper and Smith (1981) is given to illustrate the above criterion.

Hald Data

X_1	X_2	X_3	X_4	Y
7	26	6	60	78.5
1	29	15	52	74.3
11	56	8	20	104.3
11	31	8	47	87.6
7	52	6	33	95.9
11	55	9	22	109.2
3	71	17	6	102.7
1	31	22	44	72.5
2	54	18	22	93.1
21	47	4	26	115.9
1	40	23	34	83.8
11	66	9	12	113.3
10	68	8	12	109.4

High Structure Selection

Regression variable	$\text{Log}L_k$
1	-51.15
2	-49.80
3	-54.04
4	-49.06
12	-32.46
13	-52.50
14	-34.36
23	-46.13
24	-50.90
34	-40.22
123	-32.84
124	-32.56
134	-32.98
234	-35.60
1234	-34.25

The ordering of L_k is almost identical to Mallow's C_k criterion in the example above. In fact, Stone (1977) showed that C_k and the PSR Quasi-Likelihood are asymptotically equivalent in ordering subsets for the regression case (normal populations). Geisser and Eddy (1979) argued that this is true for the PSR - Bayes criterion as well. Thus, when sample size is large, say $N > 50$, Mallow's C_k may be an adequate substitute for the PSR criterion in regression problems because of the cost in computing the former is cheaper. See Daniel and Wood (1971) for details of Mallow's C_k.

3. Lower Structure Selection

When no likelihood is available for one or more of the possible models, we may use a more primitive point predictive function

$$y(Y, X, M_k) = h_k(y).$$

The deviation of the observations y_j from their predicted values $\hat{y}_j(Y_{(j)}, X_j, M_k) = h_k(Y_{(j)})$ are given by the discrepancy

$$D_k = \frac{1}{N} \sum_{j=1}^{N} d(y_j, \hat{y}_j),$$

where $d(y_j, \hat{y}_j)$ is some measure of distance between y_j and $\hat{y}_j$. The lower structure sample reuse procedure selects the model which minimizes the discrepancy D_k. Geisser (1974, 1975) introduced the general procedure. Geisser and Eddy (1979) applied it to the models of the previous section.

For comparison of various subsets of the q independent variables in

a multiple regression, the PSR squared error discrepancy for the first K is

$$D_k = N^{-1} \sum_{j=1}^{N} (y_j - X_j' b_{(j)})^2 \quad .$$

Allen (1974) termed this D_k as PRESS (Prediction Sum of Square). See also Lee (1971) for use in growth curve model selection. The set of independent variables which has the smallest discrepancy is chosen.

When this is applied to the Hald data again, the ordering of the subsets of variables induced by the criterion D_k is very similar to that induced by the residual mean square. We, however, have to be cautious before claiming they are equivalent since PRESS is not a simple function of the residual sum of squares.

Low Structure Selection Criterion

Regression Variable	D_k
1	130.74
2	92.47
3	201.26
4	91.86
12	7.22
13	170.62
14	9.32
23	53.98
24	112.45
34	22.62
123	6.92
124	6.57
134	7.27
234	11.30
1234	8.49

Lee (1971), Lee and Geisser (1975) applied the above procedures to select the best among various growth curve models.

4. The Computer Program

4.1 The Computation of L_k

Given $X_{n\times p}$, $Y_{n\times 1}$, $Y_{(j)}$ $(n-1)\times 1$, $X_{(j)}$ $(n-1)\times p$, X_{j} $1\times p$ etc., we can simplify L_k. Here, for the convenience of computing, X is X' in Section 2. Suppose $n \geq p$, there is an orthogonal matrix Q such that $Q'X = \begin{pmatrix} R \\ 0 \end{pmatrix}$, where R is upper triangular. Thus

$$X'X = X'QQ'X = (R' \; 0) \begin{pmatrix} R \\ 0 \end{pmatrix} = R'R \quad .$$

This decomposition is termed QR in Chap. 9, LINPACK (Dongarra et al., 1979). If positive diagonal elements are chosen for R, it is equivalent to Cholesky decomposition. The idea of QR decomposition applied to the subroutines SQRDC, SQRSL, SPODI of LINPACK which I used for our program. Now,

$$\text{(i)} \quad c_j = 1 - X_j (X'X)^{-1} X_j' = 1 - v_{jj} \quad .$$

Thus, we can compute $v_{jj} = ||RX_j'||$

$$\text{(ii)} \quad b_{(j)} = (X_{(j)}'X_{(j)})^{-1} X_{(j)}'Y_{(j)},$$

$$= b - (X'X)^{-1} X_j' \hat{e}_j c_j^{-1}$$

where $b = (X'X)^{-1} X'Y$ and $\hat{e}_j = y_j - \hat{y}_j$, the residual for j^{th} observation.

See Weisberg (1979) or Cook and Weisberg (1982).

Since $X_j b_{(j)} = X_j b - X_j (X'X)^{-1} X_j' \hat{e}_j c_j^{-1}$,

then $y_j - X_j b_{(j)} = \hat{e}_j + (1 - c_j) \hat{e}_j c_j^{-1}$.

Thus $c_j (y_j - X_j b_{(j)})^2 = \hat{e}_j^2 c_j^{-1} = \hat{e}_j^2 (1 - v_{jj})^{-1}$

SQRDC computes Q as the product of $m = \min\{p, n-1\}$ Householder transformation, then from SQRSL, we can get $\hat{e}_j$ plus b and many other standard output of multiple regressions.

(iii) $a_{(j)}^2 = (Y_{(j)} - X_{(j)} b_{(j)}) (Y_{(j)} - X_{(j)} b_{(j)})'$

substitute $b_{(j)}$ by (*), we get an easier form to compute:

$$= \sum_{i=1}^{N} \left(\hat{e}_i + X_i (X'X)^{-1} X_j' \hat{e}_j c_j^{-1} \right)^2$$

$$- \left(\hat{e}_j + X_j (X'X)^{-1} X_j' \hat{e}_j c_j^{-1} \right)^2$$

Since $X_j (X'X)^{-1} X_j' = 1 - c_j$, the second term becomes

$$\left(\hat{e}_j + (1 - c_j) \hat{e}_j c_j^{-1} \right)^2 = \left(\hat{e}_j c_j^{-1} \right)^2 .$$

$X_i (X'X)^{-1} X_j'$ commonly termed as v_{ij} is relatively costly to get. Again I apply SPODI to get R'R then compute v_{ij}.

4.2 The Best Model

We may select one of the three following outputs:

(i) L_k of any specified model,

(ii) Overall best model

(iii) Best model among the same number of variables.

For (ii) and (iii), the famous "Leaps and Bounds" algorithm cannot apply, for L_k is not a monotonely function of residual sum of squares. The natural sequence algorithm is used. It is for 4 variables, 1, 2, 3, 4, 12, 13, etc. For details, see Furnival and Wilson (1972) or Seber (1977). Residual mean of square and PRESS are also included in the output. The order, however, is based on L_k.

4.3 The description of the data

1. The data should include no more than 199 observations. Each observation has no more than 24 independent variables. (At this moment, the field length is designed as small as possible. Actually, it is $\leq$ 120000. We can extend to 24K, if necessary, to read a larger data set.)

2. The data is on a local file called Tape 2 with the following layout:

(i) The first line of the file may have an arbitrary identifier in the first 10 columns, the number of observations on the file, ending in column 19, then the number of columns (independent and dependent variables), ending in column 29. The word format may appear after the number of columns.

(ii) The second line of the file is a formatted statement

beginning with a left parenthesis in the first column (E, F, G, X, 0 Formats are allowed.) The rest of the file consists of the data.

Example of the input: (Hold data)

```
                                  13          5
(4(2X,F2.0),2X,F5.1)
   7   26    6   60   78.5
   1   29   15   52   74.3
  11   56    8   20  104.3
  11   31    8   47   87.6
   7   52    6   33   95.9
  11   55    9   22  109.2
   3   71   17    6  102.7
   1   31   22   44   72.5
   2   54   18   22   93.1
  21   47    4   26  115.9
   1   40   23   34   83.8
  11   66    9   12  113.3
  10   68    8   12  109.4
```

4.4 The example to access

The program is stored in the name AMATCH written in MNF language. To access the program, we need to type

```
OLD,AMATCH
/GET,TAPE2=HALD
/X,FETCH,LINPACK/V=MNF
/MNF,K,I=AMATCH
```

The terminal will print

```
=====A PREDICTIVE APPROACH TO THE BEST MODEL SELECTION.=====

NUMBER OF CASES=  13 , NUMBER OF COLUMNS=  5

 YOU HAVE 3 CHOICES,TYPE; 1 IF YOU NEED A SPECIFIED MODEL.
   ;2 IF YOU NEED THE BEST OVERALL MODEL.
   ;3 IF YOU WANT BEST MODEL AMONG  SAME NUMBER OF VARIABLES.
? 2
TYPE 0 IF NO CONSTANT,1 IF CONSTANT IN THE MODEL.
? 1
THE BEST MODEL IS  2 1
        WITH THE PSR QUASI-BAYES CRITERION        -32.4308
      THE PSR SQUARED ERROR DISCREPANCY IS          7.2217
```

If you need another selection, get the data file and run again.

```
/GET,TAPE2=HALD
/MNF,K,I=AMATCH
```

Again, the terminal shows:

```
 YOU HAVE 3 CHOICES,TYPE; 1 IF YOU NEED A SPECIFIED MODEL.
   ;2 IF YOU NEED THE BEST OVERALL MODEL.
   ;3 IF YOU WANT BEST MODEL AMONG  SAME NUMBER OF VARIABLES.
? 3
 HOW MANY VARIABLES IN YOUR BEST MODEL?
? 2
 TYPE 0 IF NO CONSTANT,1 IF CONSTANT IN THE MODEL.
? 1
THE BEST MODEL WITH   2 VARIABLES IS 2 1
   THE PSRQUASI-BAYES CRITERION          -32.4308
   THE PSR SQUARED ERROR DISCREPANCY IS      7.2217
```

You may continue if you want the best among another number of variables.

```
 DO YOU WANT OTHER BEST MODEL?(YES OR NO)
? YES
 HOW MANY VARIABLES IN YOUR BEST MODEL?
? 3
 TYPE 0 IF NO CONSTANT,1 IF CONSTANT IN THE MODEL.
? 1
THE BEST MODEL WITH   3 VARIABLES IS 4 2 1
   THE PSRQUASI-BAYES CRITERION          -32.4967
   THE PSR SQUARED ERROR DISCREPANCY IS      6.5655
 DO YOU WANT OTHER BEST MODEL?(YES OR NO)
? NO
```

For the specified model, you may have to change the order of columns for independent variables.

```
 YOU HAVE 3 CHOICES,TYPE; 1 IF YOU NEED A SPECIFIED MODEL.
   ;2 IF YOU NEED THE BEST OVERALL MODEL.
   ;3 IF YOU WANT BEST MODEL AMONG  SAME NUMBER OF VARIABLES.
? 1
TYPE COLUMS FOR X,THEN TYPE COLUMN FOR Y

? 1 2 3 4 5

TYPE K,K IS THE NUMBER OF PREDICTIVE VARIATES IN OUR SUBSET OF THE MODEL.
THEN TYPE 0 IF NO CONSTANT,1 IF CONSTANT IN THE MODEL.

? 4 1
THE MODEL IS  1 2 3 4
THE PSR BY QUASI-BAYES CRITERION IS        -34.1757
THE RESIDUAL MEAN SQUARE IS                5.9830
THE PSR SQUARED ERROR DISCREPANCY  IS      8.4882

 TYPE YES IF YOU WANT TO TRY ANY OTHER MODEL.
? YES
 TYPE YES IF YOU LIKE THE COLUMNS TO BE REARRANGED.
? NO

TYPE K,K IS THE NUMBER OF PREDICTIVE VARIATES IN OUR SUBSET OF THE MODEL.
THEN TYPE 0 IF NO CONSTANT,1 IF CONSTANT IN THE MODEL.

? 3 1
THE MODEL IS  1 2 3
THE PSR BY QUASI-BAYES CRITERION IS        -32.8100
THE RESIDUAL MEAN SQUARE IS                5.3456
THE PSR SQUARED ERROR DISCREPANCY  IS      6.9231

 TYPE YES IF YOU WANT TO TRY ANY OTHER MODEL.
? YES
```

Columns rearranged begin next page.

```
 TYPE YES IF YOU LIKE THE COLUMNS TO BE REARRANGED.
? YES
TYPE COLUMS FOR X,THEN TYPE COLUMN FOR Y

? 2 3 4 1 5

TYPE K,K IS THE NUMBER OF PREDICTIVE VARIATES IN OUR SUBSET OF THE MODEL.
THEN TYPE 0 IF NO CONSTANT,1 IF CONSTANT IN THE MODEL.

? 3 1
THE MODEL IS  2 3 4
THE PSR BY QUASI-BAYES CRITERION IS       -35.5716
THE RESIDUAL MEAN SQUARE IS                8.2016
THE PSR SQUARED ERROR DISCREPANCY  IS     11.2964

 TYPE YES IF YOU WANT TO TRY ANY OTHER MODEL.
? YES
 TYPE YES IF YOU LIKE THE COLUMNS TO BE REARRANGED.
? NO

TYPE K,K IS THE NUMBER OF PREDICTIVE VARIATES IN OUR SUBSET OF THE MODEL.
THEN TYPE 0 IF NO CONSTANT,1 IF CONSTANT IN THE MODEL.

? 2 1
THE MODEL IS  2 3
THE PSR BY QUASI-BAYES CRITERION IS       -46.2076
THE RESIDUAL MEAN SQUARE IS               41.5443
THE PSR SQUARED ERROR DISCREPANCY  IS     53.9802

 TYPE YES IF YOU WANT TO TRY ANY OTHER MODEL.
? NO
```

End of the demonstration.

4.5 The Program

```
/LIST
00100 PROGRAM PATMS(INPUT,OUTPUT,TAPE2,TAPE3)
00110 COMMON IC(25),WKAREA(25),X(200,25),RSD(200),Y(200)
00120+,N,NC,IFMT(6),IS
00130C PRINT TIME & DATE
00140C
00150 CALL SECOND(TIN)
00160 CALL CLOCK(ICLOCK)
00170 CALL DATE(IDATE)
00180 PRINT 1000,IDATE,ICLOCK
00190 1000 FORMAT(//2A10)
00200C PRINT TITLE
00210 PRINT 1010
00220 1010 FORMAT(/'=====A PREDICTIVE APPROACH TO THE BEST MODEL '
00230+,'SELECTION.====='//)
00240 READ(2,1020) N,NC,IFMT
00250 1020 FORMAT(10X,I9,I10/6A10)
00260 PRINT 1030,N,NC
00270 1030 FORMAT('NUMBER OF CASES=',I4,' , NUMBER OF COLUMNS=',I3//)
00280 PRINT,'YOU HAVE 3 CHOICES,TYPE; 1 IF YOU NEED A SPECIFIED MODEL.'
00290 PRINT,'  ;2 IF YOU NEED THE BEST OVERALL MODEL.'
00300 PRINT,'  ;3 IF YOU WANT BEST MODEL AMONG  SAME NUMBER OF ',
00310+'VARIABLES.'
00320 7 READ,ITYPE
00330 IF (ITYPE .EQ. 1) THEN
00340  CALL SPECIFY
00350 ELSE
00360 IF (ITYPE .EQ. 2)  THEN
00370   CALL ALL
00380 ELSE
00390 IF (ITYPE .EQ. 3)  THEN
00400  CALL SAME
00410 ELSE
00420 PRINT,'TYPE 1 OR 2 OR 3 ONLY. TRY AGAIN.'
00430 GO TO 7
00440 END IF
00450 END IF
00460 END IF
00470 STOP
00480 END
```

```
00490 SUBROUTINE SPECIFY
00500 COMMON IC(25),WKAREA(25),X(200,25),RSD(200),Y(200)
00510+,N,NC,IFMT(6),IS
00520 REAL DK
00530 115 PRINT 1040
00540 1040 FORMAT('TYPE COLUMS FOR X,THEN TYPE COLUMN FOR Y'/)
00550 READ,(IC(I),I=1,NC)
00560 116 PRINT 1050
00570 1050 FORMAT(/'TYPE K,K IS THE NUMBER OF PREDICTIVE VARIATES',
00580+ ' IN OUR SUBSET OF THE MODEL.'/'THEN TYPE 0 IF NO CONSTANT,'
00590+  ,'1 IF CONSTANT IN THE MODEL.'/)
00600 READ,KX,IS
00610 KC=IS+KX
00620C
00630 CALL QUASI(KX,WLK,DK,KC)
00640    RMS=0
00650    DO 1115 I=1,N
00660       RMS=RMS+RSD(I)**2
00670    1115 CONTINUE
00680 RMS=RMS/(N-KC)
00690 PRINT 1009,(IC(I),I=1,KX)
00700 1009 FORMAT('THE MODEL IS ',9I2)
00710 PRINT 1010,WLK,RMS,DK
00720 1010 FORMAT('THE PSR BY QUASI-BAYES CRITERION',
00730+ ' IS',F15.4/'THE RESIDUAL MEAN SQUARE IS',6X
00740+,F15.4,/'THE PSR SQUARED ERROR DISCREPANCY  ',
00750+ 'IS',F11.4)
00760 PRINT 1300
00770 1300 FORMAT(/' TYPE YES IF YOU WANT TO TRY ANY OTHER MODEL.')
00780  READ 1001,IANS
00790    1001 FORMAT(A3)
00800      IF (IANS .EQ. 3HYES) THEN
00810           PRINT,'TYPE YES IF YOU LIKE THE COLUMNS TO BE REARRANGED.'
00820        READ 1002,JANS
00830   1002 FORMAT(A3)
00840       IF (JANS .EQ. 3HYES) GO TO 115
00850            GO TO 116
00860         ELSE
00870          GO TO 19
00880 END IF
00890 19 RETURN
00900 END
```

```
00910 SUBROUTINE ALL
00920 INTEGER IND(25),IB(25),MIB(25)
00930 COMMON IC(25),WKAREA(25),X(200,25),RSD(200),Y(200)
00940+,N,NC,IFMT(6),IS
00950 NP=NC-1
00960 IC(NC)=NC
00970 PRINT 1050
00980 1050 FORMAT('TYPE 0 IF NO CONSTANT,1 IF CONSTANT IN THE MODEL.')
00990 READ,IS
01000 RMAX=-10000.0
01010 DO 1 L=1,NP
01020 1 IND(L)=0
01030 M=NP
01040 2 DO 3 L=M,NP
01050   IF (IND(L) .LT. L) GO TO 3
01060   IND(L-1)=IND(L-1)+1
01070  IND(L)=IND(L-1)
01080 3 CONTINUE
01090 4 IND(NP)=IND(NP)+1
01100 DO 5 L=M,NP
01110 5 IC(NP-L+1)=IND(L)
01120 MN=NP-M+1
01130 KX=MN+IS
01140 CALL QUASI(MN,WLK,DK,KX)
01150 IF (WLK .GT. RMAX) THEN
01160   RMAX=WLK
01170   DDK=DK
01180   MAX=MN
01190 DO 10 I=1,MN
01200 10 MIB(I)=IC(I)
01210 END IF
01220 IF (IND(NP) .LT. NP) GO TO 4
01230 IF (IND(M) .EQ. M) M=M-1
01240 IF (M .GT. 0) GO TO 2
01250 PRINT 110,'THE BEST MODEL IS ',(MIB(I),I=1,MAX)
01260 PRINT 120,'         WITH THE PSR QUASI-BAYES CRITERION ',RMAX
01270 PRINT 130,'       THE PSR QUARED ERROR DISCREPANCY IS ',DDK
01280 110 FORMAT(A10,A8,9I2)
01290 120 FORMAT(4A10,A3,F15.4)
01300 130 FORMAT(4A10,A2,4X,F12.4)
01310 RETURN
01320 END
```

```
01330 SUBROUTINE SAME
01340 REAL DK,DDK
01350 INTEGER IND(25),MIC(25),IB(25)
01360 COMMON IC(25),WKAREA(25),X(200,25),RSD(200),Y(200)
01370+,N,NC,IFMT(6),IS
01380 IC(NC)=NC
01390 15 PRINT,'HOW MANY VARIABLES IN YOUR BEST MODEL?'
01400 NP=NC-1
01410 READ,IVAR
01420 PRINT,'TYPE 0 IF NO CONSTANT,1 IF CONSTANT IN THE MODEL.'
01430 READ,IS
01440 KX=IVAR+IS
01450 RMAX=-10000.0
01460 DO 1 L=1,NP
01470 1 IND(L)=0
01480 M=NP
01490 2 DO 3 L=M,NP
01500   IF (IND(L) .LT. L) GO TO 3
01510   IND(L-1)=IND(L-1)+1
01520   IND(L)=IND(L-1)
01530 3 CONTINUE
01540 4 IND(NP)=IND(NP)+1
01550 DO 5 L=M,NP
01560 5 IC(NP-L+1)=IND(L)
01570 IF (IND(NP-IVAR+1) .NE. 0 .AND. IND(NP-IVAR) .EQ. 0)
01580+THEN
01590 CALL QUASI(IVAR,WLK,DK,KX)
01600 IF (WLK .GT. RMAX) THEN
01610   RMAX=WLK
01620   DDK=DK
01630   DO 10 I=1,IVAR
01640 10 MIC(I)=IC(I)
01650 END IF
01660 END IF
01670 IF (IND(NP) .LT. NP) GO TO 4
01680 IF (IND(M) .EQ. M) M=M-1
01690 IF ( M .GT. 0) GO TO 2
01700 PRINT 110,'THE BEST MODEL WITH ',IVAR,' VARIABLES',' IS',
01710+(MIC(I),I=1,IVAR)
01720 PRINT 120,'   THE PSRQUASI-BAYES CRITERION IS ',RMAX
01730 PRINT 130,'   THE PSR SQUARED ERROR DISCREPANCY IS ',DDK
01740 PRINT,'DO YOU WANT OTHER BEST MODEL?(YES OR NO)'
01750 READ 1002,JANS
01760 1002 FORMAT(A3)
01770 IF (JANS .EQ. 3HYES) GO TO 15
01780 110 FORMAT(2A10,I3,A10,A3,9I2)
01790 120 FORMAT(3A10,A2,1X,F15.4)
01800 130 FORMAT(3A10,A9,F12.4)
01810 RETURN
01820 END
```

```
01830 SUBROUTINE QUASI(IVAR,WLK,DK,KC)
01840C*
01850 INTEGER QRAUX(25),JPVAT(25)
01860 COMMON IC(25),WKAREA(25),X(200,25),RSD(200),Y(200)
01870+,N,NC,IFMT(6),IS
01880 REAL LK,DK,KK,ASQ(200),W(25),CT(25),SUM1(25),CF(25,25),
01890+ SUM2(200),DET(2),C(200,25),WORK(25),
01900+V(200)
01910C
01920C INPUT DATA
01930C
01940 REWIND 2
01950 READ(2,1051)
01960 1051 FORMAT(/)
01970 DO 1060 I=1,N
01980 READ(2,IFMT)(WKAREA(J),J=1,NC)
01990 IF (NC-1 .EQ. 0 ) GO TO 9
02000 DO 1080 J=1,IVAR
02010      X(I,J+IS)=WKAREA(IC(J))
02020      C(I,J+IS)=X(I,J+IS)
02030 1080 CONTINUE
02040 9 Y(I)=WKAREA(IC(NC))
02050  IF(IS .EQ. 0 ) GO TO 1060
02060  X(I,1)=1
02070 C(I,1)=1
02080 1060 CONTINUE
```

```
02090C
02100C USE LINPACK PACKAGE TO COMPUTE BETA ARRAY FOR REGRESSION
02110C MODEL OF THE DATA PICKED
02120C
02130C*
02140C*  COMPUTE BETA,RESIDUALS ETC.
02150C*
02160 CALL SQRDC(C,200,N,KC,QRAUX,JPVT,WORK,0)
02170 CALL SQRSL(C,200,N,KC,QRAUX,Y,WORK,RSD,B,RSD,WORK,10,INFO)
02180     RMS=0
02190     DO 1115 I=1,N
02200        RMS=RMS+RSD(I)**2
02210     1115 CONTINUE
02220 RMS=RMS/(N-KC)
02230 IF (INFO .EQ. 0 ) GO TO 13
02240 18 PRINT,' THE SOLUTION IS INCORRECT,BECAUSE THERE ARE
02250+ DEPENDENCIES IN THE COLUMNS OF X.'
02260 GO TO 19
02270 13 DO 17 I=1,N
02280  DO 221 K=1,KC
02290  221 CT(K)=X(I,K)
02300   V(I)=0.
02310 CALL STRSL(C,200,KC,CT,11,INFO)
02320 IF (INFO .NE. 0 ) GO TO 18
02330      DO 826 J=1,KC
02340        V(I)=V(I)+CT(J)**2
02350       826 CONTINUE
02360 17 CONTINUE
02370 CALL SPODI(C,200,KC,DET,11)
02380 DO 1113 I=1,KC
02390      DO 1114 J=1,KC
02400        IF (J .GE. I ) THEN
02410          CF(I,J)=C(I,J)
02420        ELSE
02430          CF(I,J)=CF(J,I)
02440        END IF
02450      1114 CONTINUE
02460 1113 CONTINUE
```

```
02470C*
02480C* COMPUTE V(I,I)=X(I)*(XTRANSPOSE*X)**(-1)*X(I)
02490C*
02500 DO 14 I=1,N
02510    DO 15 J=1,KC
02520      C(I,J)=0.
02530       DO 16 K=1,KC
02540          TEMP=X(I,K)*CF(K,J)
02550          C(I,J)=TEMP+C(I,J)
02560       16 CONTINUE
02570   15 CONTINUE
02580 14 CONTINUE
02590C* COMPUTE ASQ(I)
02600 DO 25 I=1,N
02610   SUM1(I)=0.0
02620   DO 30 J=1,N
02630     PRODUCT=0.0
02640     DO 31 K=1,KC
02650      TEMP=C(I,K)*X(J,K)
02660      PRODUCT=TEMP+PRODUCT
02670     31 CONTINUE
02680     TEMP=(RSD(J)+PRODUCT*RSD(I)/(1-V(I)))**2
02690     SUM1(I)=SUM1(I)+TEMP
02700   30 CONTINUE
02710   SUM2(I)=RSD(I)**2/(1-V(I))
02720     ASQ(I)=SUM1(I)-SUM2(I)/(1-V(I))
02730 25 CONTINUE
```

```
02740C
02750C COMPUTE TAU((N-KC)/2)
02760C
02770 KV=N-IVAR
02780 IF (KV .EQ. 1 .OR. KV .EQ. 0) THEN
02790       PRINT,' THERE ARE NOT ENOUGH OBSERVATIONS.'
02800       GO TO 19
02810 END IF
02820 K1=MOD(KV,2)
02830 KK=FLOAT(KV)
02840 IF(K1 .EQ. 0) THEN
02850       K2=KV/2-1
02860       TAU=FAC(K2)
02870       TAU1=FAC1(K2)*SQRT(3.141592653)
02880     ELSE
02890       IF (K1 .EQ. 1) THEN
02900         K2=KK/2-0.5
02910        K3=K2-1
02920         TAU=FAC1(K2)*SQRT(3.141592653)
02930         TAU1=FAC(K3)
02940       END IF
02950 END IF
02960 CONST=TAU/TAU1
02970C
02980C COMPUTE LK
02990C
03000 LK=1
03010 DK=0
03020 DO 71 J=1,N
03030       TEMP=SQRT((1-V(J))/(3.141592653*ASQ(J)))*CONST
03040+      *(1+SUM2(J)/ASQ(J))**(-KK/2)
03050       DK=DK+SUM2(J)/(1-V(J))
03060       LK=LK*TEMP
03070 71 CONTINUE
03080 DK=DK/N
03090 WLK=ALOG(LK)
03100 19 RETURN
03110 END
```

```
03120C
03130C FACTORIAL FUNCTION
03140C
03150 REAL FUNCTION FAC(N)
03160 M=N-1
03170 IF (M) 2, 3,4
03180 2 FAC=1
03190 GO TO 11
03200 3 FAC=1
03210 GO TO 11
03220 4 FAC=1
03230 DO 5 I=1,N
03240   FAC=FAC*I
03250 5 CONTINUE
03260 11 RETURN
03270 END
```

```
03280C*
03290C*  COMPUTE TAU(K+1/2)
03300C*
03310 REAL FUNCTION FAC1(N)
03320 IF (N) 2,4,4
03330 2 PRINT,'ERROR MESSAGE,THERE IS NO ENOUGH OBSERVATIONS.'
03340 GO TO 89
03350 4 FAC1=1
03360 DO 88 K=1,N
03370    TEMP=(2*K-1)
03380    FAC1=TEMP*FAC1
03390 88 CONTINUE
03400 FAC1=FAC1/(2**N)
03410 89 RETURN
03420 END
```

References

Akaike, H. (1973), "Information Theory and Extension of the Maximum Likelihood Principle," 2nd International Symposium on Information Theory, eds. Petrov and Czáki, Budapest: Akademiai Kiado, 267-281.

Allen, D.M. (1974), "The Relationship between Variable Selection and Data Augmentation and a Method for Prediction," Technometrics, 16, 125-127.

Clayton, M., Geisser, S. and Jennings, D. (1984). A Comparison of several model selection procedures. Bayesian Econometrics. P. Goel, ed. North Holland (in press).

Cook, R.D. and Weisberg, S. (1982), Chap. 2 and Appendix, Residuals and Influence in Regression, New York: Chapman and Hall.

Daniel, C. and Wood F. (1971), Chap. 6, Fitting Equations to Data, New York: Wiley.

Dongarra, J., Bunch, J.P., Moler C.B., and Stewart, G.W. (1979), Chap. 9, The LINPACK Users Guide, Philadelphia: SIAM.

Draper, N. and Smith, H. (1966, 1981) Applied Regression Analysis, New York: Wiley.

Furnival, G. and R. Wilson (1974), "Regression by Leaps and Bounds," Technometrics, 16, 499-511.

Geisser, S. (1964), "Posterior Odds for Multivariate Normal Classifications," JRSSB, 26, 69-76.

———— (1965), "Bayesian Estimation in Multivariate Analysis," AMS, 36, 150-159.

———— (1974), "A Predictive Approach to the Random Effect Model," Biometrika, 61, 101-107.

———— (1975), "The Predictive Sample Reuse Method with Applications," JASA, 70, 320-328.

———— (1983), Handouts for the course, Multivariate Analysis.

Geisser, S. and Eddy, W. (1979), "A Predictive Approach to Model Selection," JASA, 74, 153-160.

Lee, J.C. On the Generalized Growth Model, Ph.D. dissertation, Nov. 1971. SUNY at Buffalo.

Lee, J.C. and Geisser, S. (1975), "Applications of Growth Curve Prediction," Sankhyā, Ser. A, 37, 239-256.

References continued

Seber, G.A.F. (1976), Chap. 12, Linear Regression Analysis, New York: Wiley.

Stone, M. (1977), "An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion," JRSSB, 39, 44-47.

Weisberg, S. (1979), Applied Linear Regression, New York: Wiley.

———————— (1984), Lecture Notes for the course, Statistical Computation.